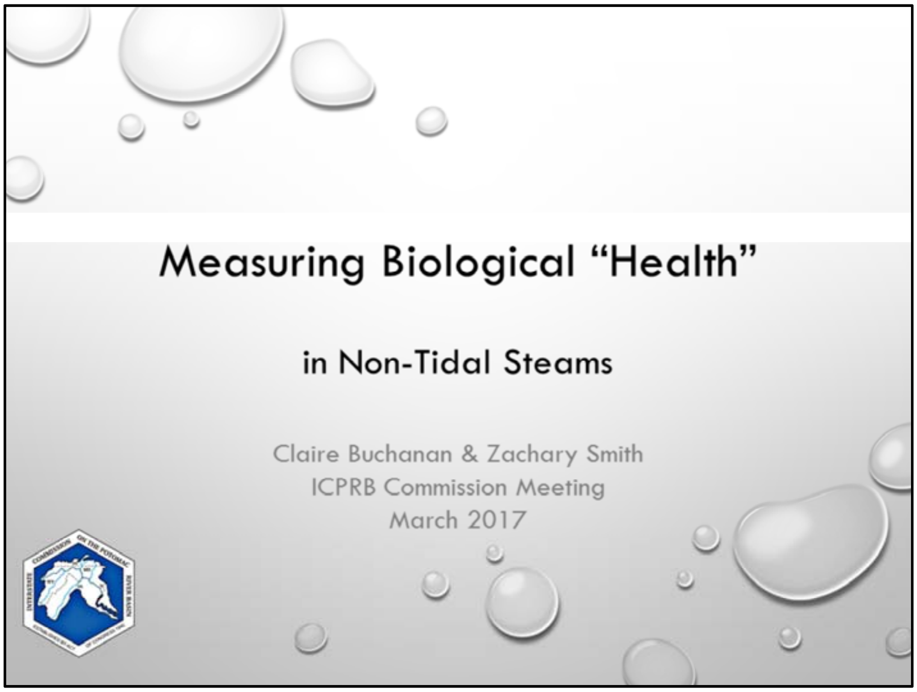
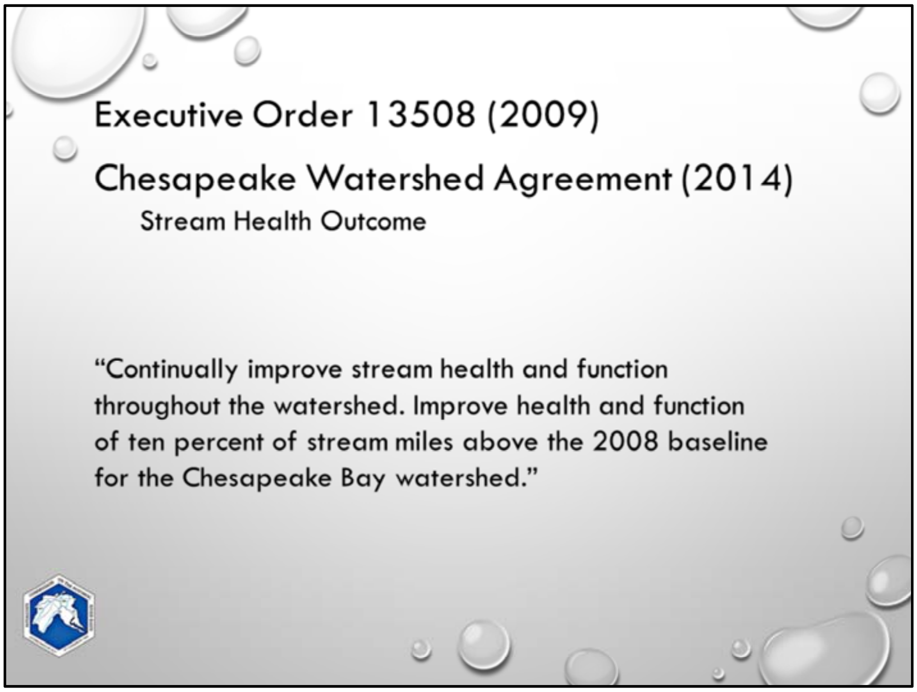
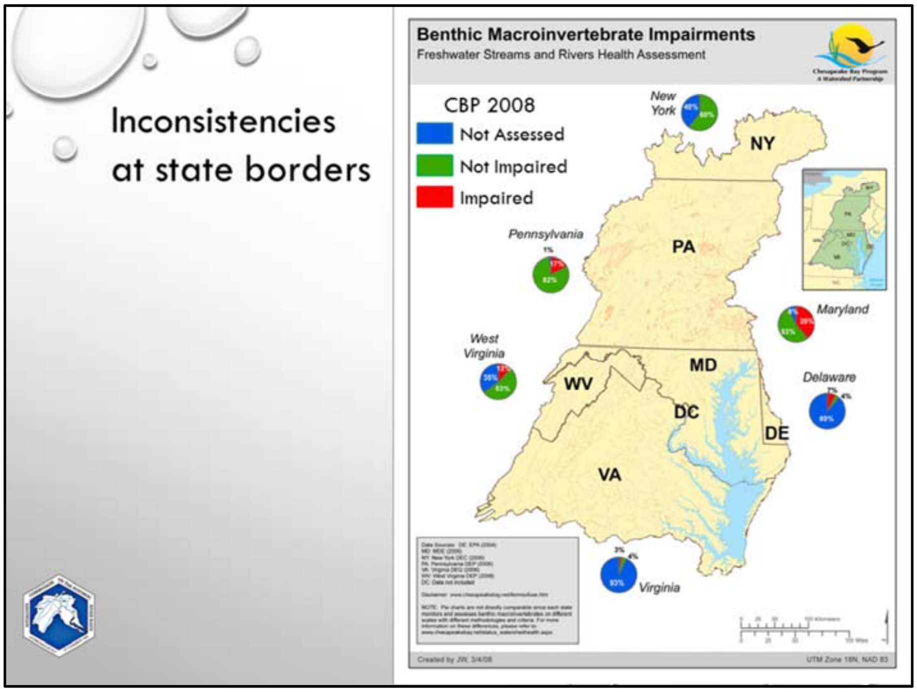


Staff introduced.



The 2009 EO and the 2014 Chesapeake Watershed Agreement specifically call for improvements in stream health and function. The reasoning behind this is that benefits from the best management practices on the land will first be seen in local streams before they are seen downstream in Chesapeake Bay. The Chesapeake Bay Program wants to document **local benefits to the biota** of non-tidal streams and rivers, in part to encourage public support for pollutant reductions. Also, in recent years, a lot of effort and funding have gone in to stream restoration projects. Needed was some quantitative measure of success.

Physical and chemical measurements capture some aspects of stream health and function. However, the biological communities living in streams are the **ultimate measure of stream health**. Stream macroinvertebrates are an obvious choice for this purpose. States have monitoring stream macroinvertebrates for a long time - in some instances since the 1970s. Field and laboratory methods are fairly consistent across states because of early EPA guidance (Rapid Bioassessment Protocols).



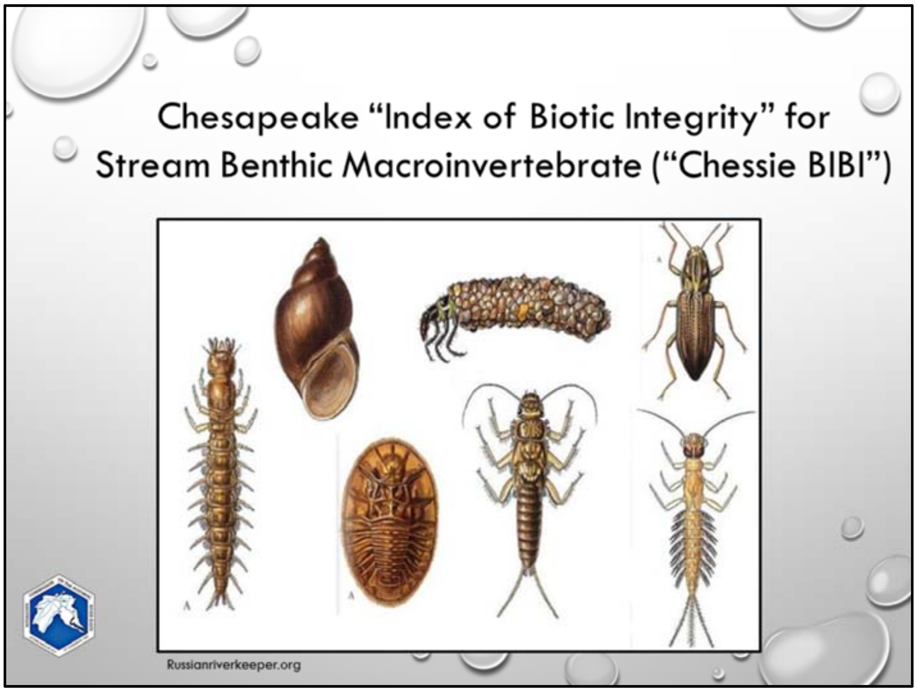
March 2008 – The CBP Non-Tidal Water Quality Workgroup created a basin-wide map of stream biological health using state-specific impairment decisions. Found big differences. e.g., Virginia 93% not assessed vs Pennsylvania 1% not assessed; New York no impairments (even in heavily agricultural Chemung River area) vs Maryland 39% impaired.

Difference not due to field and lab methods (what is measured), but rather to:

- Different thresholds and biocriteria (when is the biota impaired?)
- Different assessment units in standards (where does an impairment occur?) e.g., Maryland evaluates streams on a watershed basis, Virginia evaluates streams on a stream segment basis
- Different stressor ID methods (how is the cause of impairment identified?) e.g., state stressor identification methods are different and still evolving but have good guidance from EPA (CADDIS website)

Workgroup conclusion: results did not adequately represent stream condition on a basin-wide scale.

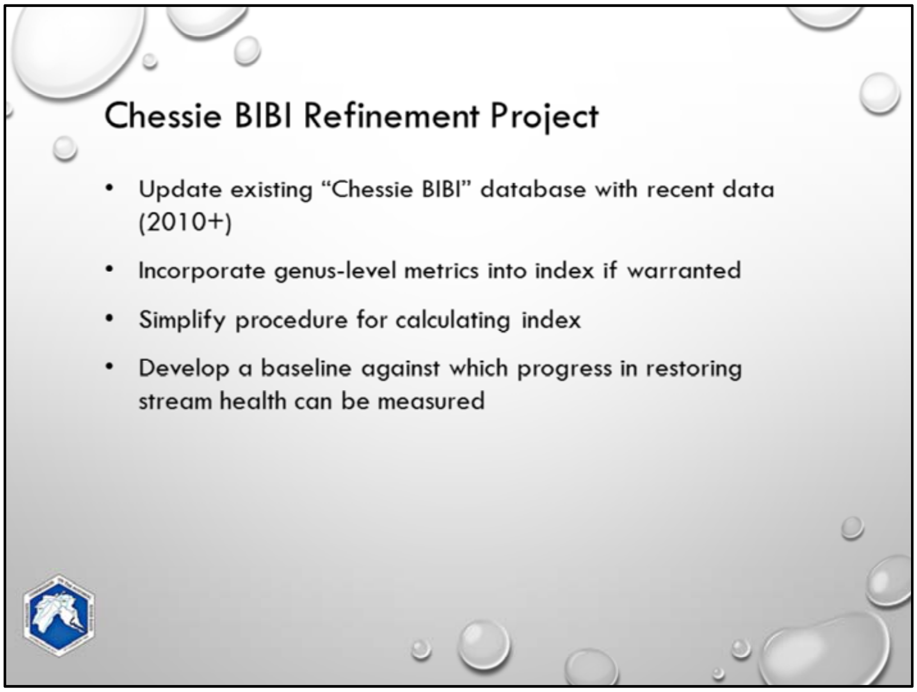
ICPRB staff LeAnne Astin had developed a Potomac basin-wide index of biotic integrity from the macroinvertebrates raw counts. These results were noticed by the workgroup (predecessor of the current Stream Health Workgroup) and ICPRB was asked to lead an effort to build a similar IBI for the entire Bay watershed.



First attempt was successful – 2011 report (Publication # ICPRB 11-01)
Zachary Smith will describe our second attempt.

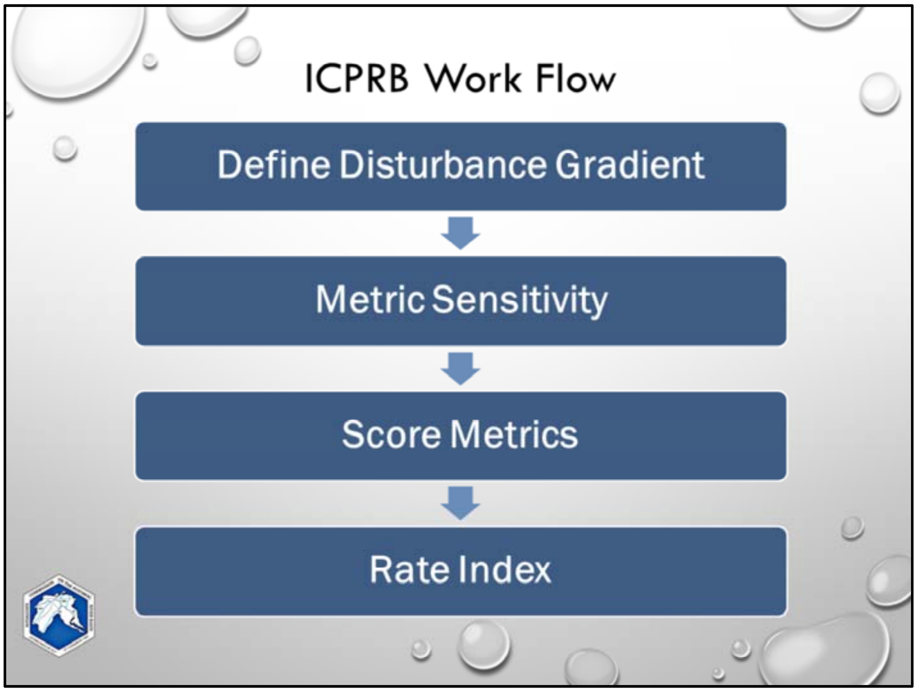
The Chessie BIBI is an Index of Biotic Integrity (IBI) developed using Stream Benthic Macroinvertebrates.
"Benthic" = organisms living on the stream bottom.
Macroinvertebrate= Invertebrates that can be observed with the naked eye. These include but are not limited to:

1) Aquatic Insects
2) Snails
3) Clams
4) Mussels
5) Crayfish
6) Freshwater Shrimp/Scuds



An update of the underlying data and a refinement of the original Chessie BIBI was begun in 2015.
Project objectives:

1) The existing "Chessie BIBI" database was updated. Now includes data ranging from 1992-2016.
2) A genus-level IBI was developed and compared to a family-level IBI.
   Note: Phylum > Class > Order > Family > Genus > Species
3) The statistical programming language, R, was used to simplify the index development process and the calculation of the IBI.
4) The ultimate goal for the Chesapeake Bay Program (CBP) is to create a 2008 baseline to measure progress of restoration efforts within the Chesapeake Bay basin. This aspect of the project is still pending and requires instruction from the CBP to proceed forward.

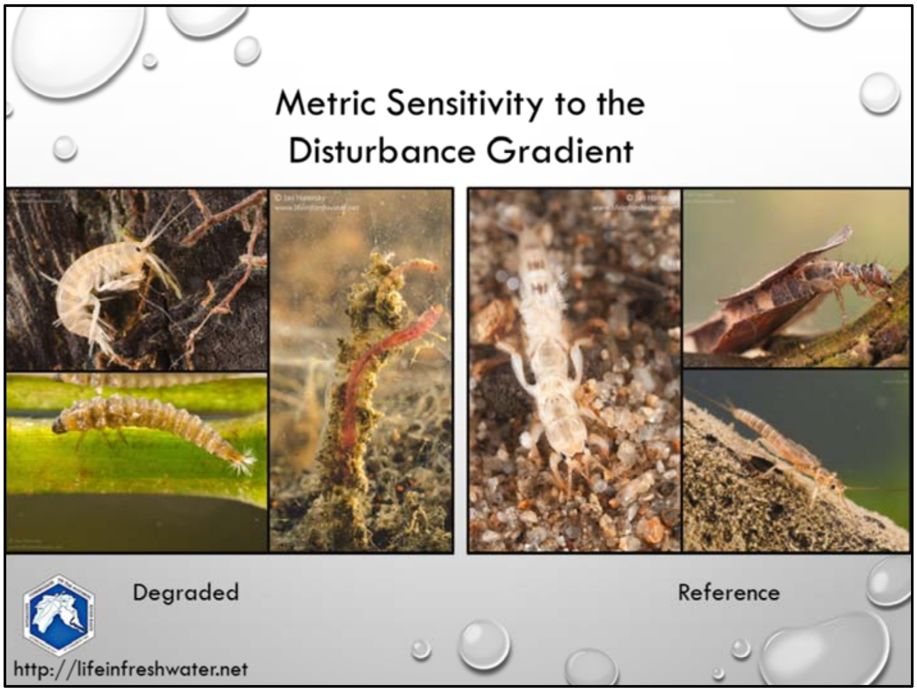


Process for creating an index of biotic integrity (IBI):

1) First define the **disturbance gradient**. During this step, biological measures are IGNORED. The disturbance gradient is defined by habitat features and water quality parameters. Essentially, we are identifying “Good” and “Bad” sites.
2) The next step requires calculation of biological measures, or “metrics.” Some examples include the percentage of predators, the number of taxa collected (richness), and percentage of mayflies. **Metric sensitivity** is an assessment of the biological metrics to identify those that best reflect the disturbance gradient. (We are identifying “Good Bugs” and “Bad Bugs.”)
3) A most sensitive metrics are selected to create the final index. However, the metric values must be scored prior to finalizing the index. **Scoring the metrics** standardizes the metric values. After the metrics have been scored, the average metric value is used as a single representative IBI value.
4) Finally, the IBI values range from 0 to 100. Zero represents a poor condition and 100 represents a good condition. However, it can be difficult to interpret the differences between an IBI value of 76 and an IBI value of 62. To make the results easier to interpret we create generalized rating categories that classify different ranges of IBI values. In a later slide we will see **five rating categories** (i.e., Very Poor, Poor, Fair, Good, and Excellent) used to represent the IBI values.

The disturbance gradient is defined using environmental variables. At this stage of the IBI development the biological measures are ignored. The disturbance gradient will be defined using two extremes, the Degraded and Reference. Degraded refers to streams with poor environmental conditions and the Reference refers to streams that have an expected natural condition.

Looking at the images above, your initial reaction may be able to classify the image on the left as a degraded stream and the image on the right as a reference stream. They most likely your initial reaction is correct but when we develop the disturbance gradient we have to have a standard methodology for classifying these sites as Degraded, Reference, or somewhere in-between. In the “Chessie BIBI,” the we defined the disturbance gradient using 8 habitat parameters (i.e., bank stability, bank vegetation, channel alteration, embeddedness, epifaunal substrate, flow, riffle/run/pool ratio, and sedimentation) and 3 water quality variables (i.e., pH, dissolved oxygen, and specific conductivity).



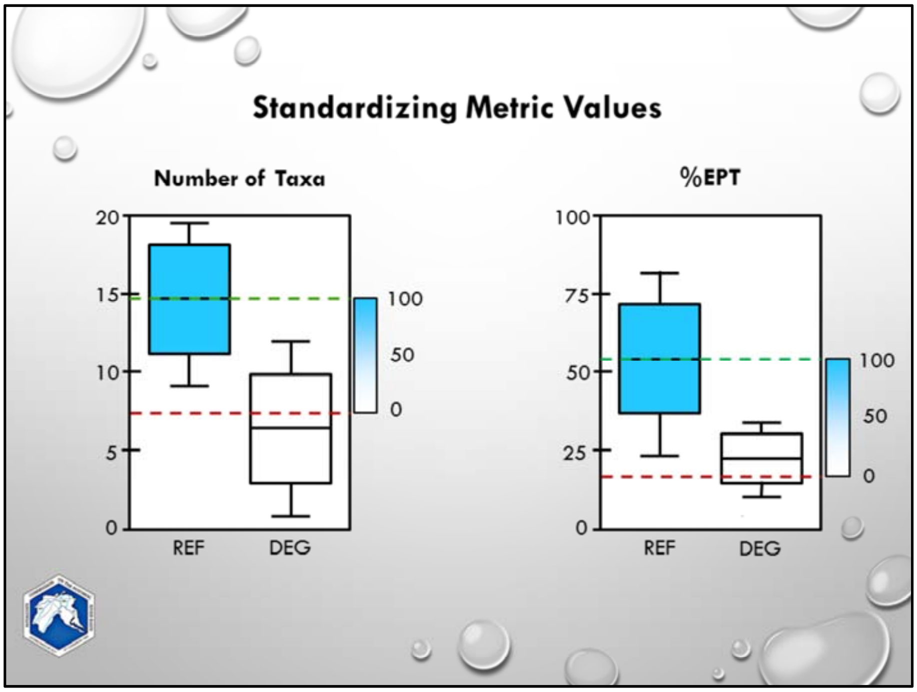
The next step is to find biological measures (i.e., metrics) that reflect the defined disturbance gradient. At the most basic level, we are looking for “Good Bugs” and “Bad Bugs.” Most IBIs select different metrics and there are no metrics that work 100% of the time. However, there are some metrics that frequently occur in IBIs developed in the U.S. The example taxa to the right represent EPT taxa. EPT metrics occur in many IBIs and can be represented as the percentage of the community comprised of EPT individuals or the richness of EPT taxa (the number of EPT taxa).

E = Ephemeroptera (Mayflies)
P = Plecoptera (Stoneflies)
T = Trichoptera (Caddisflies)

On the other end of the spectrum are organisms which tend to occur more frequently in degraded conditions. The most infamous of these taxa are the Chironomidae or non-biting midge larvae. Two of these taxa can be seen in the vertical image, above the Degraded text. This is a very diverse family of insects, many of which can tolerant a variety of degraded conditions. One aspect that makes the midges more successful in the degraded conditions, is many of these taxa develop rapidly. In extreme cases, some of these taxa can develop in as little as 2 weeks, but the majority develop over the course of several months to a year. Development refers to the life cycle of the organism. In the case of the midges, it refers to a development cycle of: egg > larvae (shown above) > pupae > adult (only terrestrial stage). Midges spend a relatively short time in the water compared to many EPT taxa, which typically spend a minimum of 1 year to several years in one of the aquatic stages. In other words, the midges get in and out of a Degraded stream quickly, while the other taxa have to deal with the stressor for a longer period of time.

Additionally, several of the midge taxa have hemoglobin in their hemolymph (i.e., blood). The reddish/pink pigmentation you see in the image above is reflected by the hemoglobin. In general terms, the hemoglobin allows the midges to handle oxygen much more efficiently than other taxa. Therefore, the midges can survive in very low dissolved oxygen conditions that might stress or kill other aquatic organisms.

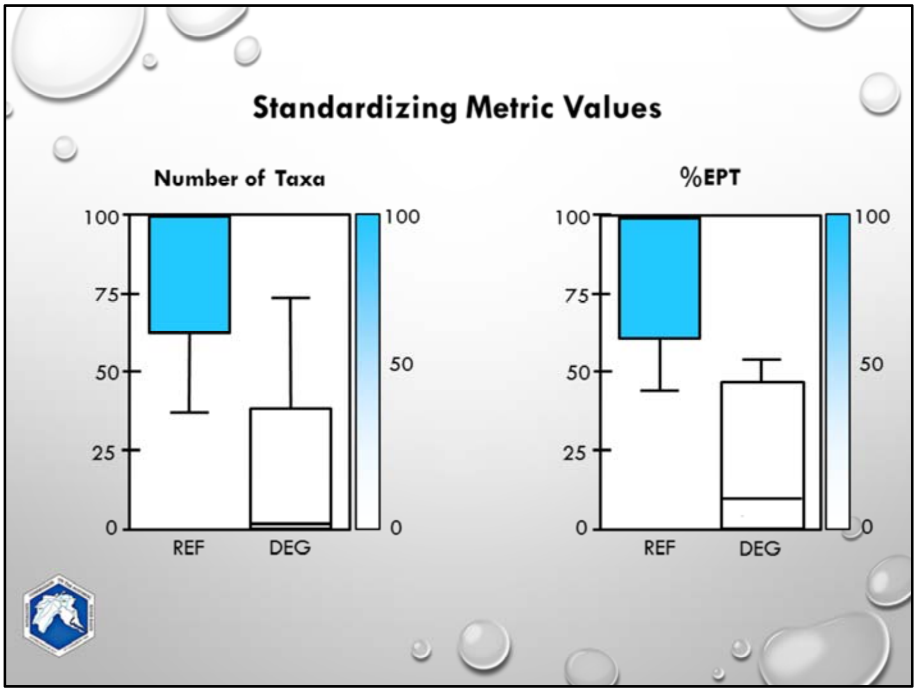
Ultimately, five to ten metrics that are sensitive to the disturbance gradient are select for the index being developed.



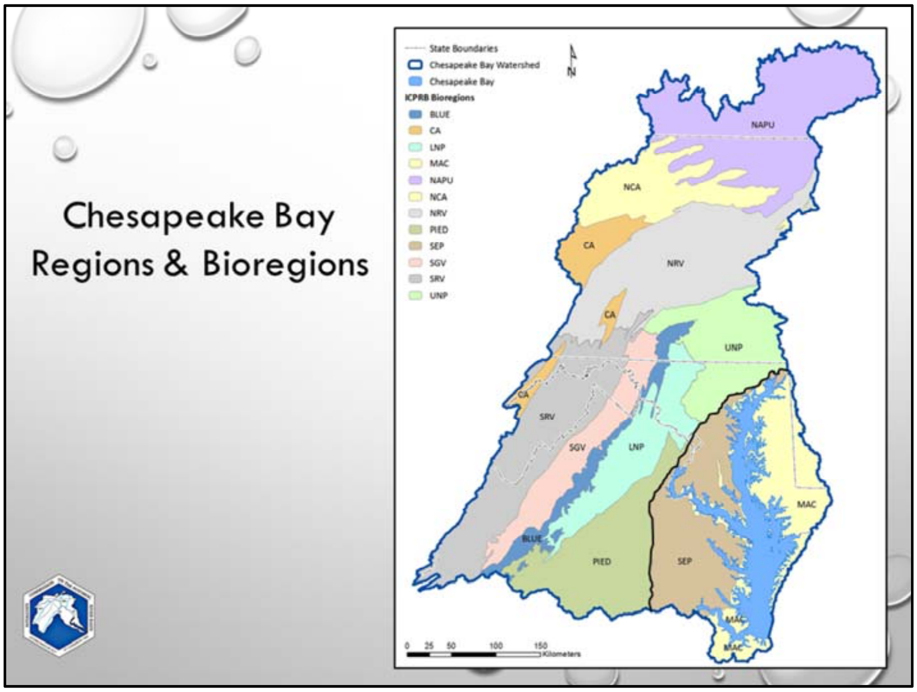
Before discussing the standardization of metric values it is important to understand the box-and-whisker plots, which we will use in several of the following slides. Box-and-whisker plots represent the distribution of data. We have grouped the data into two categories: 1) the blue boxes represent samples collected at Reference sites and 2) the white boxes represent samples collected at Degraded sites. The boxes represent the interquartile range or the central 50% of the data distribution. The lower end of the box is known as the 25th percentile. This means that 25% of the data is represented below that line and 75% of the data is represented above that line. The upper end of the box represents the 75th percentile. Therefore, 75% of the data is represented below this line and 25% of the data is represent above the line. The black line within the box represents the median or the middle of the data distribution.

For the purposes of this presentation, I just want you to notice the separation of the two boxes representing Reference and Degraded samples. I have selected two example metrics (i.e., Number of Taxa and the Percentage of EPT) to demonstrate the scoring or standardization of the metric values. Notice on the y-axis (the vertical axis) that the ranges are inconsistent between the Number of taxa (0-20) and %EPT (0-100). Averaging the values of these two metrics for a particular site would not present us with a meaningful value; the metric values must be standardized to a continuous gradient from 0-100. You may notice that the %EPT metric is already ranges from 0-100, so why do we need to score this metric? Answer: We need to treat the metrics equally. Any equation that we apply to one metric we need to apply to the other metrics.

The gradient box to the left of each of the plots is a visual representation of the scoring procedure that we are going to apply. Two thresholds must be established for ICPRB scoring procedure. All values greater than or equal to the upper threshold (green) are given a score of 100. All values less than or equal to the lower threshold (red) are given a score of 0. Values in between the threshold are assigned scores from 0-100.

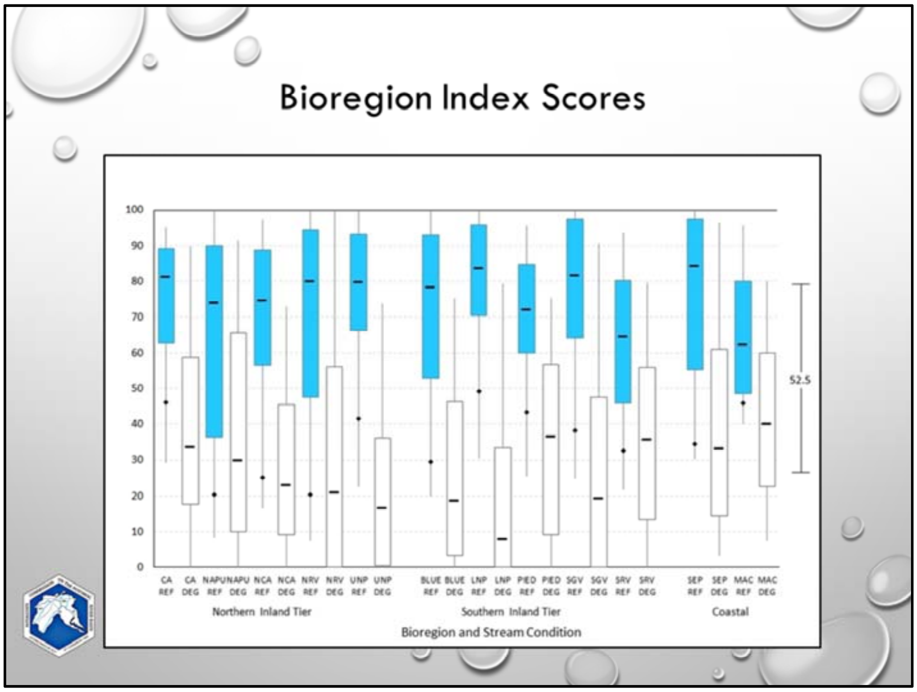


The example metrics have been scored and now both metrics have a range from 0-100.



This is a map of the Chesapeake Bay basin (watershed). Portions of the basin are represented in NY, PA, MD, DE, DC, VA, and WV.

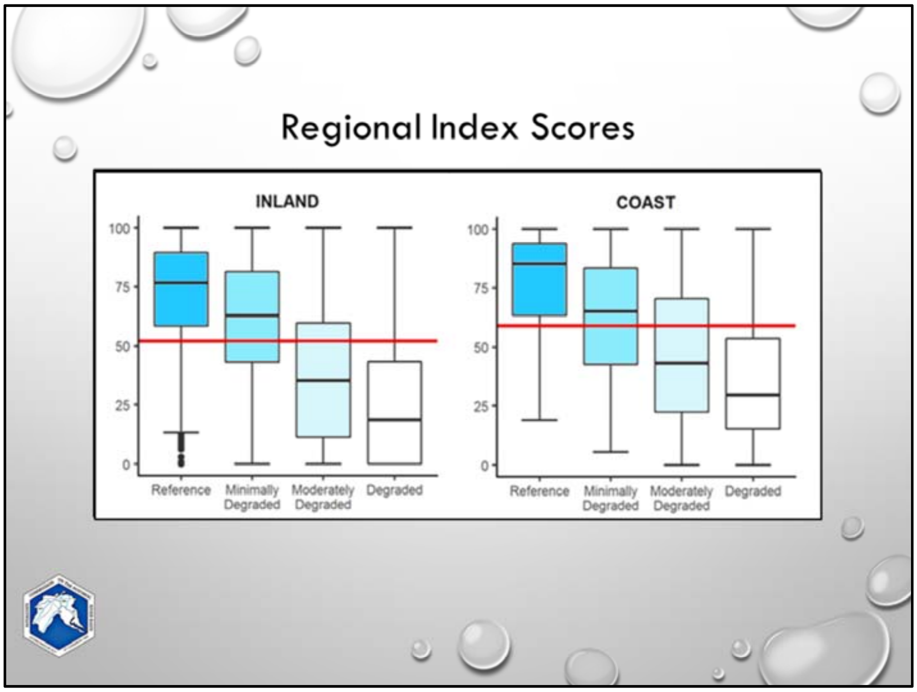
In this iteration of the "Chessie BIBI," we explored three spatial scales. 1) **Entire basin** (lowest spatial resolution). We attempted to develop a single index (i.e., one standard set of metrics) for the entire basin, and realized that the environmental conditions and taxa were very different immediately surrounding the Chesapeake Bay (Coast), relative to the remainder of the basin (Inland). 2) **Two regions**, i.e., Inland and Coast. A thick black line separates these two areas on the map. 3) **Bioregion** (highest spatial resolution). Bioregions are areas with similar topography (slope), geology (karst in Great Valley vs sandstones and shales in neighboring Ridge & Valley), weather (north vs south patterns), and connectivity of streams within river basin (e.g., Susquehanna vs Potomac/James). Bioregions may seem like an abstract concept but we recognize them when they are point out. For example, take the Blue Ridge Mountains. The dark blue polygon labeled "BLUE" on the map represents the Blue Ridge Mountain bioregion. This mountain ridge is a distinct environment from the Southern Great Valley (SGV) and the Piedmont (PIED) bioregions on either side. It is a sort of an island on the landscape and can be expected to have biological communities that differ from the neighboring bioregions. We developed an index for each of the 12 bioregions on the map.



Here are the results from the bioregion analysis. This figure again represents the data with box-and-whisker plots. The blue boxes represent Reference samples and the white boxes represent Degraded samples. Each of the 12 bioregions is represented on the x-axis (horizontal axis). The y-axis (vertical axis) represents the IBI scores, ranging from 0-100.

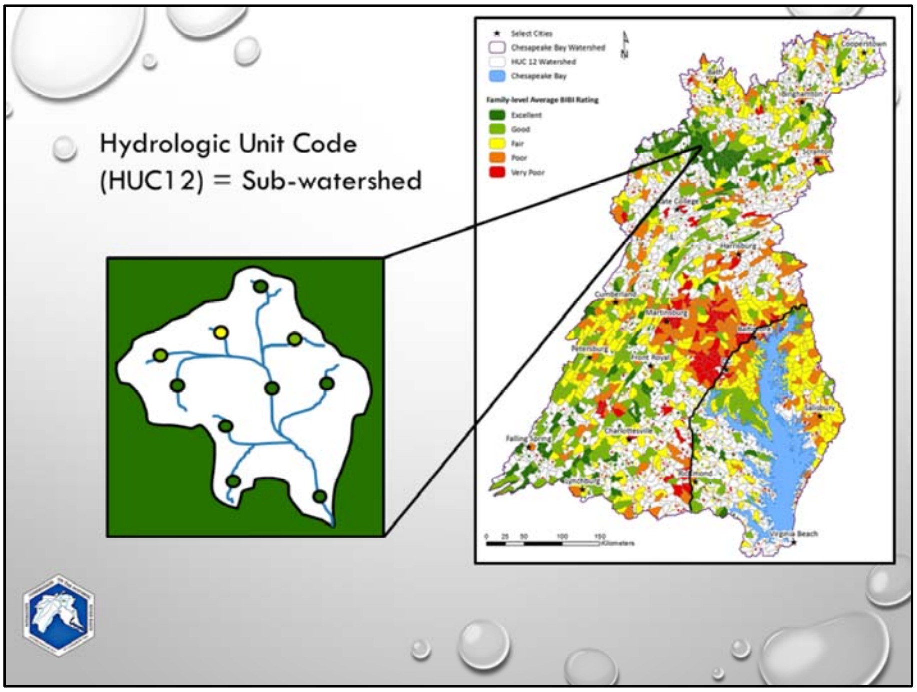
The disturbance gradient was established using the same habitat and water quality thresholds in all bioregions. Each bioregion index was developed independently with the most sensitive metrics in that bioregion. Despite the standard protocol for establishing a disturbance gradient, there is quite a bit of variability associated Reference and Degraded distributions among and between bioregions. This indicates the overall *quality* of Reference sites can differ by bioregion.

Additionally, this method is rather complex. There are 12 different indices, each with their own unique set of metrics, metric scores, and rating protocols.



Our current recommendation is to use the much more straight forward indices created for the regional spatial resolution. In the many cases these indices provided comparable, if not better, results than the bioregion indices create; Comparable results between these two spatial resolutions suggests that it is unnecessary to assess the macroinvertebrates at the bioregion level unless you really want to use an index that is fine-tuned to local features (e.g., karst, watershed slope, soils).

The figures again represent box-and whisker plots. The blue boxes represent Reference samples and the white boxes represent Degraded samples. Two additional disturbance gradient categories were added to the plot to further exemplify the downward trend from Reference to Degraded sample distributions. The red horizontal lines point to the indicate the separation between the Reference and Degraded boxes (interquartile ranges). The complete separation of the interquartile ranges means that greater than 75% (Classification Efficiency) of the time these indices are correctly identifying Reference samples as Reference samples and Degraded samples as Degraded samples.



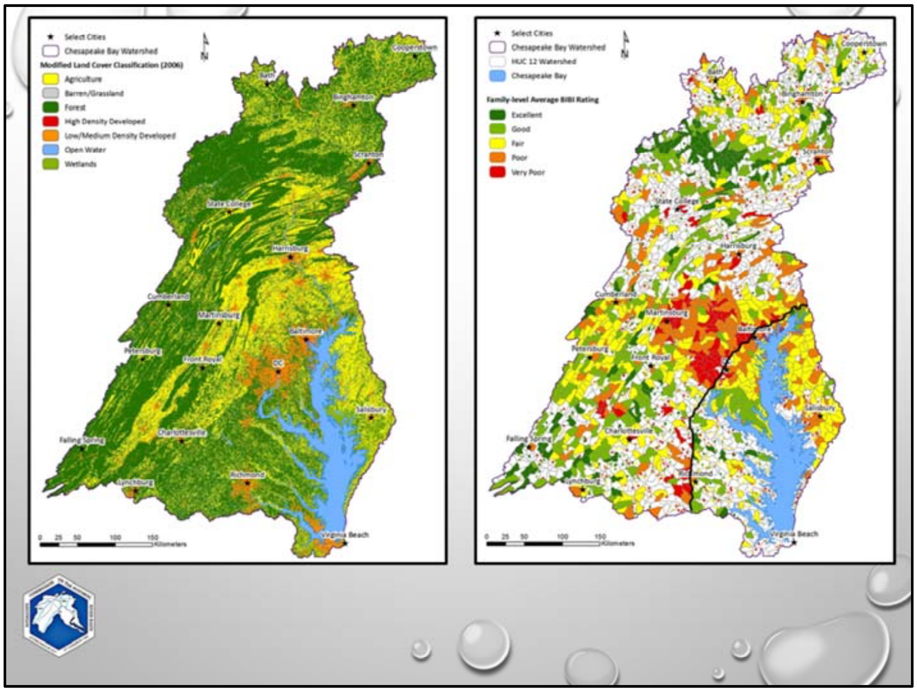
The Regional indices from the previous slide were rated to make the index values, ranging from 0-100, readily informative.

Five ratings were assigned:

1) Excellent (Dark Green)
2) Good (Light Green)
3) Fair (Yellow)
4) Poor (Orange)
5) Very Poor (Red)

The white spaces represent areas of the basin that have not been sufficiently sampled to assign a Chessie BIBI rating. The colored spaces are polygons represent Hydrologic Units, or “HUC12s,” known as sub-watersheds. The entire Chesapeake basin can be divided into HUC12s.

For a HUC12 to receive a color, ten or more samples had to have been collected within the HUC12. The index values from the ten or more samples were averaged together and the corresponding rating was assigned. The image on the left represents a single HUC12 from northern PA. Ratings for ten sampling locations within the HUC12 are presented. In this example HUC12, there are 7 Excellent samples, 2 Good samples, and 1 Fair sample. When the index values were averaged together for this HUC12, the final rating was Excellent (Dark Green).

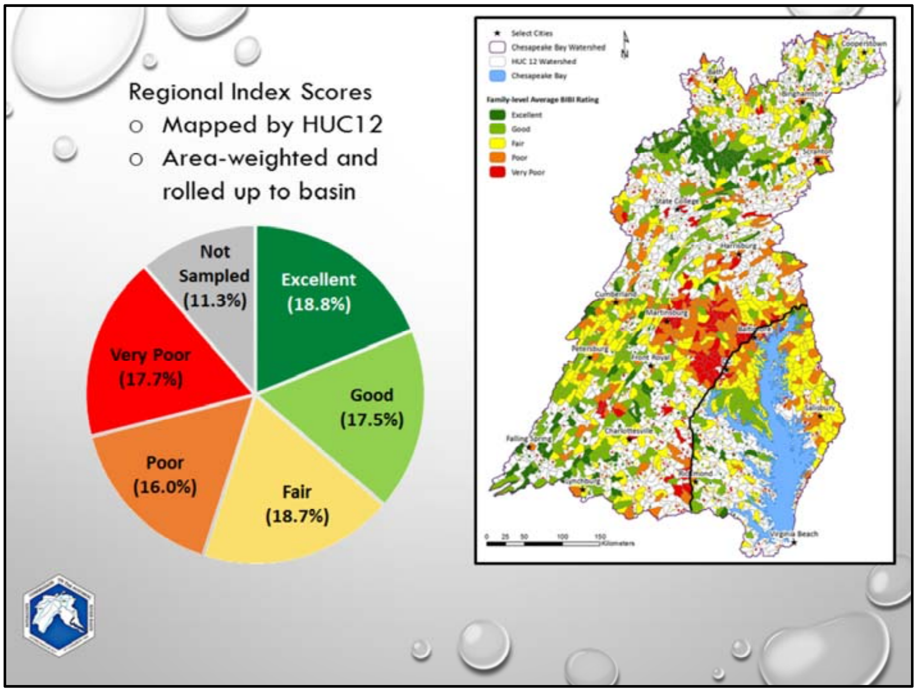


The map on the left represents land cover in the Chesapeake Bay basin (NLCD-2006).

Agriculture = Yellow
Barren = Gray
Forest = Dark Green
High Development = Red
Low Development = Orange
Open Water = Blue
Wetlands = Light Green

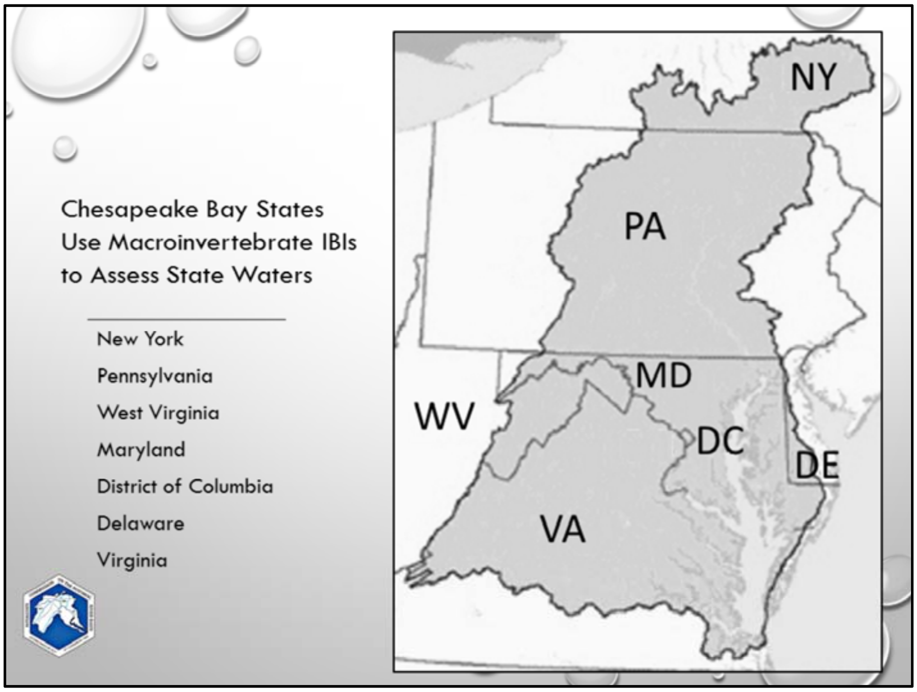
Comparing the Chessie BIBI ratings to the land-use, we can see many areas in agreement. For example, areas with urban development in the land-use map are represented by Poor (Orange) and Very Poor (Red) in the Chessie BIBI rating map. You can also see similarities between the two maps in the forested areas, such as in Northern PA and the lower portion of the Blue Ridge Mountains.

We do, however, find heavily forested HUC12's with Poor biological ratings and highly urban/agricultural HUC12's with Good biological ratings. Land use information can sometimes indicate why these ratings diverge from expected levels (e.g., metal and sediment TMDLs and acid mine drainage in heavily forested sub-watersheds; parklands and riparian buffers in urban areas).



The Chesapeake Bay Program (CBP) wants to understand the distribution of Chessie BIBI ratings within the basin. Previously, each HUC12 rating was treated equally. However, this causes issues because the HUC12s do not represent equal portions of the basin. Developed areas are more frequently sampled than remote areas, and simply adding the counts of each rating creates a result biased toward Poor. To alleviate this issue we are recommending that the CBP area weight the HUC12s. Therefore, the ratings of larger HUC12s will have proportionally more influence than the ratings of smaller HUC12s.

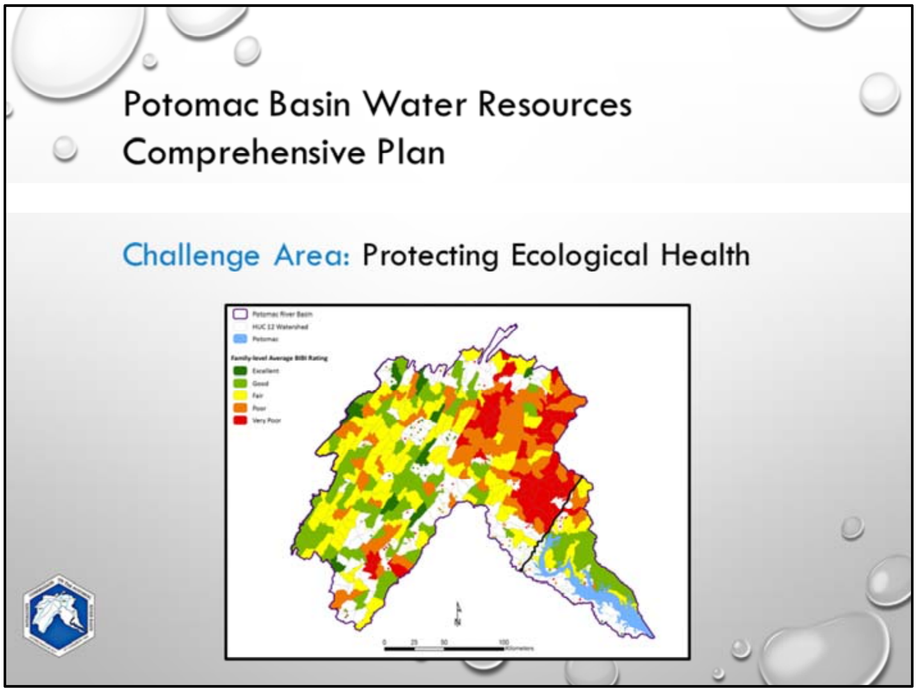
The pie chart represents the proportion of each rating represented in the Chesapeake Bay Basin using the family-level regional indices of biotic integrity.



All six states and DC use stream macroinvertebrate IBIs to assess waters. **The Chessie BIBI index is not meant to supersede state assessments of impaired waters.** These are state regulatory tools; the Chessie BIBI is not a regulatory tool.

Everyone recognizes that waters flow across state boundaries and a common measure of stream health is need for other, non-regulatory reasons such as:

- report biological "health" by waterbody and watershed
- identify and protect high quality waters
- identify degraded waters best suited for remediation
- report progress on a Chesapeake Basin level



ICPRB would like to use information on biological indicators and IBIs in the **Ecological Health chapter** of the Potomac Basin Comprehensive Water Resources Plan led by Heidi Moltz.
(Map is based on the Regional, Family-level BIBI index.)


Questions? Comments?